

2<sup>nd</sup> AMENDMENT AFTER FINAL  
August 31, 2005

JP920000257US1  
Serial No. 09/943,341

**AMENDMENTS TO THE CLAIMS**

This listing of claims will replace all prior versions, and listings, of claims in the application:

**Listing of Claims:**

1. (cancelled)
2. (currently amended) An automatic method according to claim 23 ~~[[15]]~~, wherein the step of computing the aggregated word score for said each word comprises:  
computing a score for said each word according to linguistic salience of said each word to a user profile.
3. (currently amended) An automatic method according to claim 23 ~~[[15]]~~, wherein the step of computing the aggregated word score for said each word comprises:  
computing a score for said each word according to similarities among said each word, a query and a provided topic.
4. (currently amended) An automatic method according to claim 23 ~~[[15]]~~, wherein the step of computing the aggregated word score for said each word comprises:  
computing a score for said each word according to similarities among said each word and terms in titles of the documents.
5. (currently amended) An automatic method according to claim 23 ~~[[15]]~~, wherein the step of computing the aggregated word score for said each word comprises:  
computing a score for said each word according to a ratio of an occurrence number for said each word in a document to a total occurrence number for said each word in the set of documents.

2<sup>nd</sup> AMENDMENT AFTER FINAL  
August 31, 2005

JP920000257US1  
Serial No. 09/943,341

6. (currently amended) An automatic method according to claim 23 ~~[[15]]~~, wherein the step of computing the aggregated word score for said each word comprises:

computing a score for said each word according to a ratio of a number of documents including said each word to a total number of documents in the set of documents.

7 – 8 (canceled)

9. (currently amended) A computer program product for automatically generating summaries according to claim 24 ~~[[19]]~~, wherein the computer program code means for computing the aggregated word score for said each word comprises:

computer program code means for computing a score for said each word according to linguistic salience of said each word to a user profile.

10. (currently amended) A computer program product for automatically generating summaries according to claim 24 ~~[[19]]~~, wherein the computer program code means for computing the aggregated word score for said each word comprises:

computer program code means for computing a score for said each word according to similarities among said each word, a query and a provided topic.

11. (currently amended) A computer program product for automatically generating summaries according to claim 24 ~~[[19]]~~, wherein the computer program code means for computing the aggregated word score for said each word comprises:

computer program code means for computing a score for said each word according to similarities among said each word and terms in titles of the documents.

12. (currently amended) A computer program product for automatically generating summaries according to claim 24 ~~[[19]]~~, wherein the computer program code means for computing the aggregated word score for said each word comprises:

2<sup>nd</sup> AMENDMENT AFTER FINAL  
August 31, 2005

JP920000257US1  
Serial No. 09/943,341

computer program code means for computing a score for said each word according to a ratio of an occurrence number for said each word in a document to a total occurrence number for said each word in the set of documents.

13. (currently amended) A computer program product for automatically generating summaries according to claim 24 [[19]], wherein the computer program code means for computing the aggregated word score for said each word comprises:

computer program code means for computing a score for said each word according to a ratio of a number of documents including said each word to a total number of documents in the set of documents.

14 – 15 (currently canceled)

16. (currently amended) An automatic method according to claim 23 [[15]], wherein document discourse analysis comprises identifying titles, sections, lists, paragraph boundaries and sentence boundaries of the documents.

17. (currently canceled)

18. (currently amended) An automatic method according to claim 23 [[17]], wherein said aggregate sentence score further has a weighted relationship with each of said aggregated word score, sentence position (position(s, d)) and similarity (similarity(s, S)) of the form  $SCORE[s] = \lambda_7 * \sum(SCORE[w], s \in w) + \lambda_8 * position(s, d) + \lambda_9 * similarity(s, S)$ .

19. (currently canceled)

20. (currently amended) A computer program product for automatically generating summaries according to claim 24 [[19]], wherein computer program code means for generating a set of sentences for a set of documents by document discourse analysis

2<sup>nd</sup> AMENDMENT AFTER FINAL  
August 31, 2005

JP920000257US1  
Serial No. 09/943,341

comprises computer program code means for identifying titles, sections, lists, paragraph boundaries and sentence boundaries of the documents.

21. (cancelled)

22. (previously presented) A computer program product for automatically generating summaries according to claim 24 [[21]], wherein said aggregate sentence score further has a weighted relationship with each of said aggregated word score, sentence position (position(s, d)) and similarity (similarity(s, S)) of the form

$$\text{SCORE}[s] = \lambda_7 * \sum (\text{SCORE}[w], s \in w) + \lambda_8 * \text{position}(s, d) + \lambda_9 * \text{similarity}(s, S).$$

23. (new) An automatic method for generating summaries for text documents, comprising steps of:

generating a set of sentences for a set of documents by document discourse analysis and a set of words by morphologic process;

initializing a word score for each word in the set of words, a sentence score for each sentence in the set of sentences and a score sum;

computing an aggregated word score for said each word according to an aggregate of sentence scores of sentences containing said each word and to a degree of correlation between said each word and user related information;

wherein said aggregated word score (SCORE[w]) has a weighted ( $\lambda$ ) relationship with each of said aggregated sentence score (SCORE[s]), linguistic salience of said each word to a user profile (salience(w, user summarization profile)), similarities among said each word, a query and a provided topic (salience(w, user's query or topic)), similarities among said each word and terms in titles of the documents (salience(w, title words)), a ratio of an occurrence number for said each word in a document to a total occurrence number for said each word in the set of documents

(FREQUENCY(w/d)/FREQUENCY(w/D)), and a ratio of a number of documents

2<sup>nd</sup> AMENDMENT AFTER FINAL  
August 31, 2005

JP920000257US1  
Serial No. 09/943,341

including said each word to a total number of documents in the set of documents  
(NUMBER(d, d<sub>3</sub>w)/ NUMBER(D)), of the form

$$\begin{aligned} \text{SCORE}[w] = & \lambda_1 * \text{salience}(w, \text{user summarization profile}) \\ & + \lambda_2 * \text{salience}(w, \text{user's query or topic}) \\ & + \lambda_3 * \sum (\text{SCORE}[s], s \in \omega) \\ & + \lambda_4 * \text{salience}(w, \text{title words}). \\ & + \lambda_5 * \text{FREQUENCY}(w/d) / \text{FREQUENCY}(w/D) \\ & + \lambda_6 * \text{NUMBER}(d, d_3w) / \text{NUMBER}(D); \end{aligned}$$

computing an aggregated sentence score for said each sentence according to an  
aggregate of word scores composing said each sentence and a respective sentence  
position in a section and a paragraph;

comparing an aggregate sum with said score sum, said aggregate sum being a sum  
of aggregated word scores and aggregated sentence scores; and

if said aggregate sum is different than said score sum, returning to the step of  
computing the aggregated word score; otherwise,

outputting top-ranked sentences according to sentence score as a summary of the  
set of documents, top-ranked words according to word score as a keywords list of the set  
of documents.

24. (new) A computer program product for automatically generating summaries for text  
documents, said computer program product comprising a computer usable medium  
having computer readable program code thereon, said computer readable program code  
comprising:

computer program code means for generating a set of sentences for a set of  
documents by document discourse analysis and a set of words by morphologic process;

computer program code means for initializing a word score for each word in the  
set of words, a sentence score for each sentence in the set of sentences and a score sum;

2<sup>nd</sup> AMENDMENT AFTER FINAL  
August 31, 2005

JP920000257US1  
Serial No. 09/943,341

computer program code means for computing an aggregated word score for said each word according to an aggregate of sentence scores of sentences containing said each word and computing a degree of correlation between said each word and user related information;

computer program code means for computing an aggregated sentence score for each sentence in the set of sentences according to an aggregate of word scores composing it and a respective sentence position in a section and a paragraph;

wherein said aggregated word score (SCORE[w]) has a weighted ( $\lambda$ ) relationship with each of said aggregated sentence score (SCORE[s]), linguistic salience of said each word to a user profile (salience(w, user summarization profile)), similarities among said each word, a query and a provided topic (salience(w, user's query or topic)), similarities among said each word and terms in titles of the documents (salience(w, title words)), a ratio of an occurrence number for said each word in a document to a total occurrence number for said each word in the set of documents (FREQUENCY(w/d)/FREQUENCY(w/D)), and a ratio of a number of documents including said each word to a total number of documents in the set of documents (NUMBER(d, d<sub>w</sub>)/NUMBER(D)), of the form

SCORE[w]= $\lambda_1$ \*salience(w, user summarization profile)

+ $\lambda_2$ \*salience(w, user's query or topic)

+ $\lambda_3$ \* $\sum$ (SCORE[s], s<sub>w</sub>)

+ $\lambda_4$ \*salience(w, title words)

+ $\lambda_5$ \*FREQUENCY(w/d)/FREQUENCY(w/D)

+ $\lambda_6$ \*NUMBER(d, d<sub>w</sub>)/NUMBER(D)

computer program code means for computing an aggregate sum from aggregated word scores and aggregated sentence scores;

computer program code means for determining if said aggregate sum is different than said score sum and for selectively replacing said score sum with said aggregate sum,

2<sup>nd</sup> AMENDMENT AFTER FINAL  
August 31, 2005

JP920000257US1  
Serial No. 09/943,341

each said word score with a corresponding said aggregated word score and each said sentence score with a corresponding said aggregated sentence score; and

computer program code means for outputting top-ranked sentences according to sentence score as a summary of the set of documents, top-ranked words according to word score as a keywords list of the set of documents.